DOCUMENT RESUME

ED 051 264 TM 000 573

AUTHOR	Schroder, Carole; Crawford, Patricia
TITLE	School Achievement as Measured by Teacher Ratings and Standardized Achievement Tests.
INSTITUTION	Toronto Board of Education (Ontario). Research Dept.
PUB DATE	[70]
NOTE	60p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Performance, Achievement Rating, Achievement Tests, *Comparative Analysis, Correlation, Creativity, Grade 3, Grade 6, Intelligence Quotient, Longitudinal Studies, Predictive Ability (Testing), *Predictor Variables, *Rating Scales, *Standardized Tests, Student Adjustment, *Student Evaluation, Teacher Attitudes
IDENTIFIERS	Metropolitan Achievement Test, *Teachers Rating Questionnaire

ABSTRACT

The relationships between teacher ratings and standardized achievement tests as measures of student progress are investigated. Initial teacher ratings of pupils in grade 3 and subsequently in grade 6, were contrasted to scores obtained on the Metropolitan Achievement Test (MAT). It was found that a teacher's rating does not provide the same information about the pupil's level of achievement as do the MAT scores. Although a relationship between the two measures was found, the degree of relationship depended upon what areas the teachers had rated. An examination of the data revealed a consistent pattern in the relationship between various subsections of the Teacher Rating Questionnaire and the MAT, the discrepancy scores, and the I.Q. scores. Teacher ratings of performance were much more closely related to these measures than were their ratings of pupil adjustment and creativity. Despite differences among teachers and among classes, as well as the develcpment in the pupils' characteristics from grade 3 to grade 6, there was a reasonable level of agreement among teachers' ratings of student performance. The data supports the conclusion that teachers' ratings of academic achievement are an important and dependable adjunct to the use of standardized achievement tests. The development of the instrument is described briefly and copies of two versions are appended. (Author/AG)

ED051264

SCHOOL ACHIEVEMENT AS MEASURED BY TEACHER RATINGS AND STANDARDIZED ACHIEVEMENT TESTS

Carole Schroder
Patricia Crawford

RESEARCH SERVICE

issued by the

Research Department

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TM 000 573

THE BOARD OF EDUCATION FOR THE CITY OF TORONTO

ED051264

SCHOOL ACHIEVEMENT AS MEASURED BY TEACHER RATINGS AND STANDARDIZED ACHIEVEMENT TESTS

Carole Schroder
Patricia Crawford

TABLE OF CONTENTS

	Page Number
INTRODUCTION AND REVIEW OF LITERATURE	1
PROCEDURE	7
Pupil Population	7
Description of the Measures Used	7
ORGANIZATION OF THE RESULTS	14
RESULTS	16
The Distribution of Teacher Rating Scores, MAT Scores and Discrepancy Scores	16
The Relationship Between Teacher Rating Scores and Average MAT Scores	18
Prediction of Grade Six Teacher Ratings from Grade Three Teacher Ratings	20
Comparison of Teacher Rating Scores and Discrepancy Scores	22
The Relationship Between I.Q. Scores and Average MAT Scores and Teacher Rating Scores	27
MAT Scores, I.Q. Scores and Discrepancy Scores as Combined Predictors of Teacher Ratings	28
SUMMARY AND IMPLICATIONS	32
REFERENCES	34
APPENDIX	35

FOREWORD

Teachers' ratings, in one form or another, help form the basis of many decisions made about a pupil's progress. The relationships between such ratings and standardized achievement tests will therefore be of interest to educators. Although any study in which these relationships are examined will be "technical," the following report attempts to present that information so that it is meaningful to a person who has little familiarity with statistics.

INTRODUCTION AND REVIEW OF LITERATURE

How valid are teachers' judgements of their pupils' school success? Teachers are asked on many occasions both formal and informal, by principals and parents, to evaluate and assess the progress of the pupils in their classrooms. Since "achievement" within the school system may be interpreted as being closely related to the interaction between teacher and pupil, it could be suggested that the teacher's rating of a pupil provides the most appropriate measure of the pupil's achievements. However, a pupil's performance on a standardized test of achievement often is considered to be the best indicator of his achievements.

What is the relationship between these two measures of achievement? When a teacher is asked to make a rating or judgement of a particular pupil, she will usually rate him relative to the performance and behaviour of the other pupils in the class or in the school. Consequently, an average pupil in a class consisting primarily of below average pupils may be rated as "outstanding," because relative to the other pupils in the class he is an above average student. On the other hand, if this same pupil were in a class of primarily above average pupils, his performance might be rated as "below average" relative to that of his classmates. On a standardized test, however, each pupil's scores are compared to a common referent, i.e. the norms developed for that particular test. Such norms are based on the performance of a large number of pupils in a large number of different classrooms. Does the fact that the teacher uses her own class as a "measuring stick" against which

to assess a particular pupil result in a rating which is inconsistent with that pupil's performance on a standardized test?

Although there has been some work done to investigate the relationship between teachers' ratings and standardized test measures, it has not been particularly extensive or detailed.

Finley (1966) compared pupil performance on three subtests, i.e. language arts, arithmetic reasoning and reading comprehension, of three different achievement tests[1] with teacher ratings obtained when the pupils were in grade three and grade five. The teachers were asked to indicate whether, in general, each pupil in their class was performing at, below, or above the grade level. She found some disagreement among the different standardized tests, i.e. pupils obtained higher scores on the California Achievement Test (CAT) than on either the Metropolitan Achievement Tests (MAT) or the Iowa Test of Basic Skills (ITBS), as well as some disagreement between the standardized tests and the teacher ratings. The results of the series of comparisons between each of the standardized tests and the teacher ratings are as follows:

(1) California Achievement Test: In both grade three and grade five, teacher rating scores were lower than scores obtained on all three subtests of the CAT.

(2) Metropolitan Achievement Tests: In grade three, the teacher ratings were the same as the MAT scores on two subtests, language arts and reading comprehension, but lower than the MAT arithmetic subtest scores. In grade five, the results were exactly the opposite, i.e. the teacher rating scores were lower than the MAT scores on the language arts and reading comprehension subtests and the same as the MAT arithmetic subtest scores.

---

1 California Achievement Test, Metropolitan Achievement Tests and the Iowa Test of Basic Skills.

(3) Iowa Test of Basic Skills: With one exception, the results of the comparison between grade three and grade five teacher ratings and the ITBS were identical to those discussed above for the MAT.

These differences in the extent of the agreement between teachers' ratings and achievement scores, depending upon both the particular achievement test employed and the grade of the pupil, suggest that caution is required in interpreting any such future comparisons. It should not be concluded that teachers' ratings of school performance, because they do not agree with achievement test scores, are poor estimates of the pupils' accomplishments. In fact, the differences may arise in part because the teachers' ratings allow for the inclusion of aspects of performance that are obvious to the teacher, but are not measured by the pencil and paper achievement tests. These differences may also arise in part because the achievement tests are based on nationally-standardized norms whereas the teachers' ratings are largely based on the norms that exist for that one class.

In another study (Ebbesen, 1968), teachers were asked near the end of the school year to put their kindergarten pupils into rank order in terms of how far they thought each pupil would go in school. At the end of the first, second and third grades respectively, these pupils were again placed in rank order; however, in these instances, the ranking was done on the basis of the MAT scores obtained for that year. To measure the degree of correspondence between the teacher ratings and the MAT scores, three correlation coefficients were calculated, one for each grade. Remembering that the maximum positive value of a correlation coefficient is +1.00, representing a perfect correspondence between two sets of scores,

the correlations obtained were: 0.60 in grade one; 0.54 in grade two; 0.52 in grade three. The similarity of the three coefficients indicates that a knowledge of the teacher's ranking of the pupils in kindergarten provides as good a prediction of their MAT score in grade one as it does in grade three. Although Ebbesen felt that the correlations were quite low and indicated a fairly large amount of disagreement between teacher and achievement test rankings, it must be remembered that there were few opportunities available for kindergarten teachers to observe their pupils performing the sorts of tasks that in later grades were to form the basis for the achievement test evaluations. In addition, the rankings in kindergarten were made on the basis of only <u>one</u> criterion, a prediction of future educational success. In light of these two points, Ebbesen's findings seem very encouraging.

In another study of kindergarten pupils (Meyers, Attwell and Orpet, 1968), teachers were asked to rate the overt behaviour exhibited by their pupils during the administration of a series of achievement tests. The behaviours rated included: amount of motor activity, manual dexterity, attention and effort displayed. When the pupils were in grade five, they were given the California Achievement Test. The results of this test were compared with the achievement test results and behaviour ratings obtained when the pupils were in kindergarten. It was found that one of the behaviour ratings, attention, was almost as good a predictor of the grade five achievement test scores as any of the kindergarten achievement tests, i.e. the maximum correlation between any of the kindergarten subsets and the grade five scores was 0.50 whereas the correlation between attention and the grade five achievement test scores was 0.40.

Although the three studies just discussed have approached the problem in somewhat different ways using different standardized achievement tests and pupils at a number of grade levels, they do provide some evidence for a relationship between teacher ratings and standardized test measures.

It was suggested in the introduction that teachers may base their judgements on a relative as opposed to an absolute scale, i.e. their judgements of pupil performance are related to the expectations and standards they have established for their own class, such standards, of course, varying from class to class. The following studies provide some support for this suggestion. Although Hubbard and Flesher (1953) gave their teachers criteria for judging I.Q., they found that teacher judgements were highly influenced by the particular range of intelligence test scores of the pupils in their class. This finding was confirmed by Alexander (1953), who found that teacher judgements of the extent to which a pupil was working up to his capacity were influenced by the norms that the teacher had established for his class. The experimenters observed that although the teachers were able to rank their pupils according to their ability or level of achievements, the numerical values that they gave as estimates of I.Q. frequently were consistently too high or too low for any given class.

Haberman and Raths (1968) conducted a more direct investigation of the criteria used by teachers in selecting high- or low-achieving pupils. Each of a number of sixth-grade teachers was given a fictitious list of pupils and their grade point averages. Each teacher was given a different list; some lists had a wide range of grade point averages, while others had a narrower range restricted to those scores falling

toward the centre of the distribution of possible scores. They were asked to select from the list the pupils that they thought would benefit from either enrichment or remedial classes. The results showed that the teachers selected the same number of pupils from their list irrespective of the range of grade point averages. This would seem to indicate that they were selecting pupils according to their standing relative to the rest of the class, not according to the pupil's absolute score.

On the basis of the above findings, it may be stated that teachers' estimates of their pupils' abilities and accomplishments differ somewhat from the estimates obtained using standardized tests. The sample of pupils included in each of these studies was usually relatively small, i.e. 100 or less, and the teachers asked to rate their pupils in terms of only one attribute or characteristic. Since data collected as a part of the longitudinal Study of Achievement initiated by the Research Department in 1960-1961 included scores for a large number of pupils on the following measures:

(1) Teacher Rating Questionnaires administered in grade three and grade six;

(2) Metropolitan Achievement Test administered in grade three;

(3) Otis Quick-Scoring Mental Ability Test (new edition -- Alpha Short Form) administered in grade two;

it was possible to examine in some detail the relationship between teachers' ratings of a pupil's school success and standardized test measures of his school achievement.

PROCEDURE

Pupil Population

The data on which this paper is based were collected from 520 pupils who were a part of the basic population of 8,695 in the longitudinal Study of Achievement. There are two reasons for the small number of pupils included in this study relative to the larger basic population:

(1) Although during the first five years of the Study data were collected from the entire sample of pupils available, in grade six, the teacher ratings were obtained only for those 721 pupils whose mothers had been interviewed during the previous year, i.e., 1967, when the pupils were in grade five[2].

(2) Of those 721 pupils, I.Q. scores, MAT scores, and grade three teacher ratings were available for 520 pupils; grade six teacher ratings were available for only 429 of these 520 pupils, as some had left the system or were absent during testing.

Description of the Measures Used

The Otis Quick-Scoring Mental Ability Test (new edition -- Alpha Short Form) is a standardized test of intelligence which was administered during grade two.

The Teacher Rating Questionnaires, for grade three and grade six were developed by the Research Department. They consist of four subsections, i.e. Adjustment, Performance, Creativity and Prediction of School Success, each containing a number of separate but related items. The teacher is asked to give each pupil a rating of either 0, 2, 4, 6 or 8, depending upon the extent to which the pupil displays the characteristic in question. A brief descriptive statement accompanies each possible rating and provides the teacher with a guideline for determining which

---

2 For a discussion and more detailed information concerning the interview data, the reader is referred to School Achievement: A Preliminary Look at the Effects of the Home, Research Department, 1970.

value is most appropriate for any given pupil. The following item from the grade six Adjustment subsection is given as an illustrative example:

General Adjustment Evaluation

Considering all aspects of the child's adjustment to the classroom environment, evaluate his position.

Rate 0 -- Quite out-of-place.

Rate 2 -- Seems uncomfortable.

Rate 4 -- Adequately adjusted.

Rate 6 -- Makes consistent, conscientious efforts to improve himself in relation to his school world.

Rate 8 -- Makes an effort to positively influence his classroom world.

The complete questionnaires for both grade three and grade six as well as a brief history detailing the development of the Teacher Rating Questionnaire may be found in Appendix A. It should be noted that some modifications were made on individual items in the questionnaire between grades three and six. These changes were based on experiences with the questionnaire in grade three.

The standardized achievement test employed was the Metropolitan Achievement Tests, Elementary Battery, administered in grade three. The areas covered by this test series include word knowledge, word discrimination, reading, spelling, arithmetic computation, arithmetic problem solving, language usage, punctuation and capitalization. Each pupil received a score representing his performance on each of the above subtests; these subtest scores are not combined during the scoring of the test to give a total score representing an individual pupil's achievement level. For the purposes of this study, however, average MAT scores were calculated for each pupil.

One means of evaluating a pupil's school progress is by looking at his actual achievements relative to his predicted achievements. A Discrepancy Score was devised to measure the extent to which each pupil's actual achievements corresponded to what it might be expected his achievements should be. On the basis of this score, the pupils were categorized as either under-, average-, or over-achievers. A brief digression is necessary at this point to explain the rationale and technique for determining this score.

In order to make precise predictions about any pupil's school achievements, accurate quantitative measures of his actual and potential achievements are required. Typically, it is expected that pupils of higher levels of measured intelligence (I.Q.) will do better in school than their colleagues with lower I.Q. scores. Therefore, in some instances I.Q. scores and achievement test scores have been directly compared, i.e. pupils whose achievement test score was less than their I.Q. score were identified as under-achievers, and pupils whose achievement test score exceeded their I.Q. score were labelled as over-achievers. However, serious problems associated with the use of this simple comparison procedure led Thorndike (1963) to develop a more refined technique for identifying over- and under-achievers. Thorndike's method significantly reduces the effects of one of the sources of error associated with any kind of testing, that of random or chance error. For example, whenever a pupil is given a test of any sort, his final score will likely be influenced by a number of factors in addition to the knowledge that he is asked to communicate in answering the test questions. For example, if he were not feeling well on the day of the test, his score might be lower than it would be on another occasion. If he correctly guessed some

of the answers, his score would be higher than it should be. If, when taking an I.Q. test a pupil correctly guessed some of the answers and received a high score, and took a standardized achievement test on a day when he was not feeling well and received a low score, he would probably be categorized as an under-achiever, whereas, if he had taken both tests under ideal conditions, he might have been categorized as an average-achiever. Thorndike's (1963) statistical technique for determining a predicted achievement score reduces by a considerable amount the chances of such erroneous classifications.

A second feature of Thorndike's technique is that it allows the establishment of several groups which have different discrepancy scores, i.e. under- or over-achievers, but all of which have the same average I.Q. score. Consequently, in each group there are equal numbers of pupils with high, average or low I.Q. scores, i.e. each I.Q. score is equally represented in each of the groups.

Description of Thorndike's Statistical Technique for Determining a Predicted Achievement Score

In order to determine an individual's discrepancy score, a knowledge of both his predicted and actual achievement scores is necessary. Thorndike's formula for determining the predicted achievement score, in this case, the predicted Metropolitan Achievement Test score, is as follows:

$$\text{Predicted MAT score} = \text{Individual's I.Q. score X correlation between the set of I.Q. and actual MAT scores for the group}^{3}$$

Expressed verbally, this formula means that each pupil's predicted achievement score is determined by multiplying his I.Q. score by the correlation

3 It should be noted that the successful application of this technique requires that I.Q. and achievement test scores be available for a fairly large group of pupils. A class of thirty is suggested as the minimum.

obtained between the I.Q. scores and actual achievement scores for the group being studied. The reader will remember that a correlation is always obtained between two sets of scores, and is a measure of the extent to which, on the average, an individual's score on one test can be predicted from his score on another test. This correlation coefficient functions as a correction factor for random error (as discussed above) that may exist in the data. When the predicted MAT scores have been obtained, the difference between the predicted and actual MAT scores can be calculated for each pupil and a decision made as to whether each pupil falls into the category of under-, average-, or over-achiever. If the pupil's predicted MAT score (what he should be getting, given his I.Q.) is higher than his actual MAT score by at least a certain amount, he is designated an under-achiever. A reversal of this relationship identifies the over-achiever.

Calculation of the Discrepancy Scores for the Pupils in the Study of Achievement

Before calculating predicted MAT scores, the actual MAT scores and the I.Q. scores were converted to standard scores. This is a means of making scores obtained by different testing procedures mathematically comparable. Essentially, the process involves taking each individual's score on a test, subtracting it from the group average, and dividing the result by the standard deviation[4]. Using these standard score forms of the actual MAT scores and the I.Q. scores, predicted MAT scores were obtained using the method described in the preceding section. The difference between the predicted and actual MAT scores was obtained and the pupils

4 The standard deviation is an index of the amount of variation in the test scores. By dividing each person's difference score by this index, each of the original sets of test scores is transformed so that all the sets have equal amounts of variation in them and also have the same average score, thereby making the data easier to work with. The transformation in no way changes the relationships among the scores within each individual set, or between the two sets of scores.

sorted into five categories in terms of the following criteria:

(1) If the difference between the predicted and actual MAT scores was greater than ± 1.00 the pupils were classified as either extreme over- or under-achievers, according to the direction of the difference;

(2) If the difference was between ± 0.50 and ± 0.99, the pupils were classified as either over- or under-achievers;

(3) If the difference between predicted and actual MAT scores was less than ± 0.49, the pupils were classified as average-achievers. The number of pupils falling into each of these groups is shown in Table 1. These numbers are also expressed as a percentage of the total population of 520.

TABLE 1

NUMBER AND PER CENT OF GRADE THREE PUPILS IN EACH DISCREPANCY SCORE GROUP, ALONG WITH THEIR AVERAGE I.Q. AND MAT SCORES

Achievement Group	No.	Per Cent of Total	Mean I.Q.	Mean MAT Score	
				Verbal	Arithmetic
1 - Extreme Under-Achiever	55	10.57	104.38	20.85	16.14
2 - Under-Achiever	83	15.96	104.76	23.98	19.04
3 - Average-Achiever	245	47.11	106.57	29.29	22.14
4 - Over-Achiever	91	17.50	105.91	32.86	25.48
5 - Extreme Over-Achiever	46	8.84	103.33	35.35	27.74

It should be noted that although the average I.Q. score for each group is approximately the same, i.e. about 105, the average scores on the combined verbal or the combined arithmetic subsections of the MAT increase as the actual MAT score increasingly exceeds the predicted MAT score, i.e. as a progression is made from Group 1 through to Group 5. Thus there are now available for study five groups of pupils who are virtually identical in one respect, i.e. average I.Q. score, but who differ in another respect, their discrepancy score, or the extent to which their actual achievement scores correspond to their predicted achievement scores.

ORGANIZATION OF THE RESULTS

As outlined in the previous section, the following data were available for the pupils included in this study:

(1) I.Q. score;

(2) Teacher Ratings in grade three;

(3) Teacher Ratings in grade six;

(4) Metropolitan Achievement Test scores obtained in grade three;

(5) Discrepancy Score, which could range from 1 to 5, where a score of 1 designated an extreme under-achiever and a score of 5 designated an extreme over-achiever.

These data were used in the following comparisons to examine in detail the different measures used to assess pupil progress in school:

(1) The teacher ratings in both grade three and grade six were examined to see to what extent teachers avail themselves of the full range of possible scores when evaluating pupils. Do teachers tend not to use extreme scores and show a preference for "average" ratings, or do they tend to avoid "average" ratings? To what extent are teacher rating scores distributed in the same manner as scores on standardized achievement tests?

(2) Average MAT scores were compared with teacher ratings made in grade three to examine the degree of agreement between teacher ratings and standardized tests completed in the same year.

(3) Average MAT scores were compared with teacher ratings made in grade six to give some indication of the accuracy of long-range predictions based on a knowledge of the average MAT scores.

(4) A comparison of the grade three and grade six teacher ratings was made to determine the extent to which ratings made in later grades, i.e. grade six, could be predicted on the basis of earlier ones, i.e. grade three.

(5) Teacher ratings for both grade three and grade six were compared with the discrepancy scores to establish the extent to which teachers distinguished between over-, average-, and under-achievers in their ratings on pupils' adjustment, performance, creativity, and prediction of school success.

(6) I.Q. scores were compared with average MAT scores. I.Q. scores were compared with teacher rating scores in order to determine whether the relationship between I.Q. and teacher ratings was as strong as the relationship between I.Q. and MAT scores.

(7) MAT scores, I.Q. scores, and discrepancy scores were combined in a multiple regression analysis in order to determine the optimal combination of these three different measures for the prediction of teacher ratings.

RESULTS

The Distribution of Teacher Rating Scores, MAT Scores and Discrepancy Scores

The first step in the analysis was to determine whether each of the achievement measures was normally distributed across the range of possible scores. Each of the distributions was divided at three mathematically equivalent points into five sections, and the percentage of the total number of scores falling into each section was calculated. These data are presented graphically in Figure 1. It is evident that all three measures are distributed in a similar manner, i.e. the largest percentage of the scores fall in the middle of the distribution, with approximately equal percentages falling at either end. Although this general pattern is exemplified by each of the measures, the distribution for both the MAT scores and the discrepancy scores is more accentuated than that for the teacher rating (TR) scores, i.e. a larger percentage of the MAT and discrepancy scores fall at the centre of the distribution (44 and 47% respectively) as compared with the TR scores (32%). The teachers did avail themselves of the full range of possible ratings when assessing their pupils and in addition, they tended to use the "average" rating just about as often as the ratings at either extreme.

When TR scores were compared with the standardized test scores, more pupils were rated as average on the basis of the MAT results, (i.e. 44%) than on the basis of the TR results (i.e. 32%). One implication of this finding is that teachers can discriminate better among their pupils than can standardized tests. It must be remembered, however, that the Teacher Rating Questionnaires tap a wider range of classroom behaviours

than do standardized tests and so allow the teachers to form a more distinctive description of their pupils.

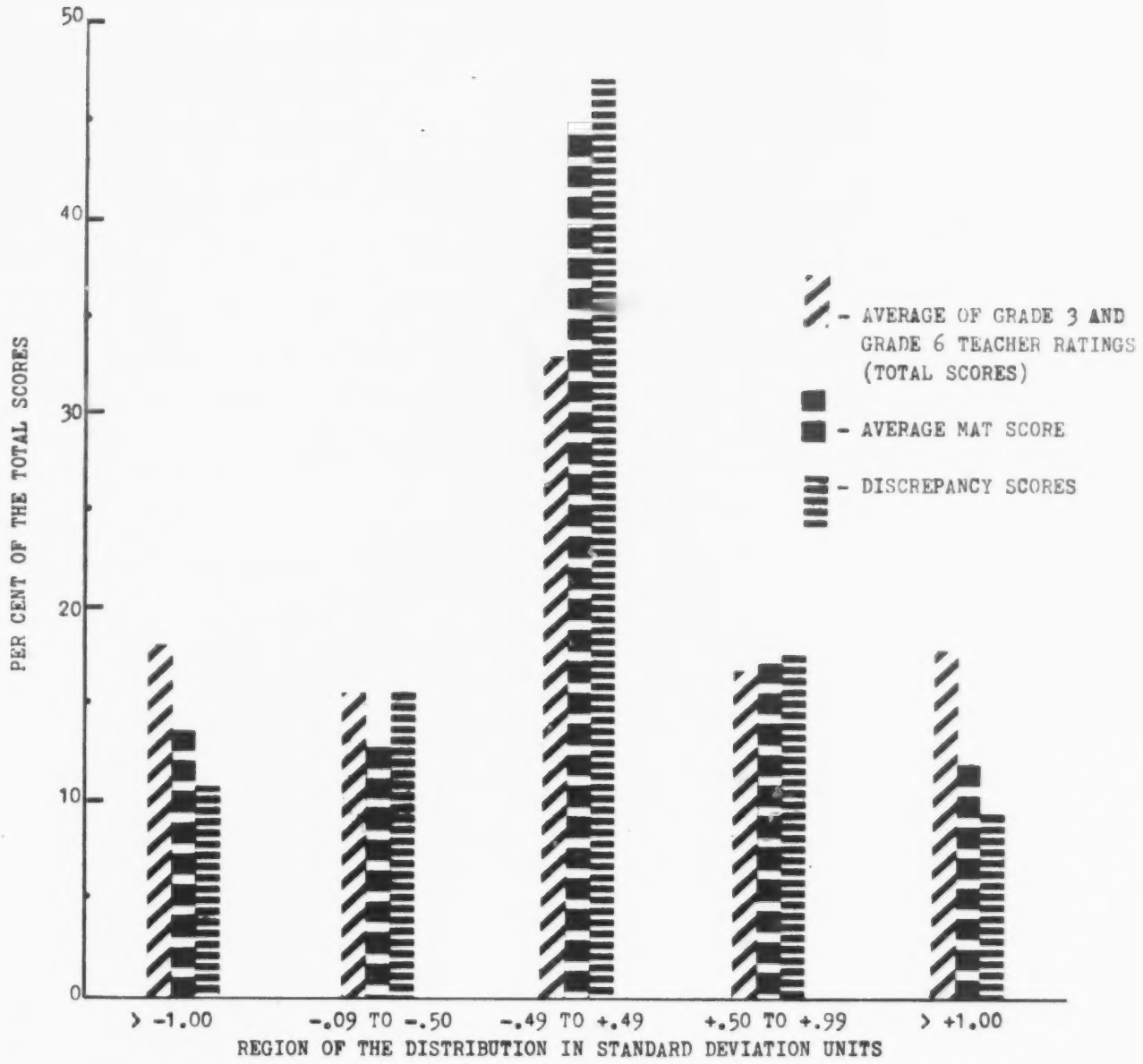

Fig. 1. Distributions of three achievement measures: teacher ratings, mean MAT scores and discrepancy scores.

The Relationship Between Teacher Rating Scores and Average MAT Scores

The scores for each subsection of each of the grade three and the grade six teacher ratings were correlated with the average MAT scores (see Table 2).

TABLE 2

CORRELATIONS OF AVERAGE MAT SCORES WITH TEACHER RATING SUBTEST SCORES

	Adjustment	Performance	Creativity	Prediction	Total
Grade 3					
r	.50	.69	.50	.67	.66
r^{2*}	.25	.47	.25	.45	.44
Grade 6					
r	.42	.62	.45	.65	.61
r^2	.18	.38	.21	.42	.37

* To obtain the per cent of variation accounted for by "r", multiply r^2 by 100.

Inspection of Table 2 revealed that in both grade three and grade six there was a stronger relationship between the average MAT scores and the scores on the performance (r = .69 and .62 for grades three and six respectively) and prediction (r = .67; .65) sections than between the average MAT scores and the adjustment (r = .50; .42) and creativity (r = .50; .45) sections of the Teacher Rating Questionnaire. Since the MAT taps primarily academic performance, it was not surprising that there should be a stronger relationship with those sections of the teacher ratings related to academic performance as compared with those sections concerned with other aspects of the pupils' classroom behaviour.

It should be noted that while the absolute values of the MAT-TR correlations ranged from .42 to .69, the two sections most relevant to standardized test scores, performance and prediction, maintained values in excess of .60 over the three year interval. A correlation of .60 indicates that 36% of the variability can be explained and, while modest, is quite respectable when maintained over time.

One further point must be made regarding the relationship between the teacher ratings and the average MAT scores. Although, as described above, the same pattern of results was obtained for both grade three and grade six, the values of the correlations obtained between the grade six teacher ratings and the mean MAT scores were slightly lower than those obtained with the grade three teacher ratings. There are at least two factors which could account for the small decrease in the value of these correlation coefficients:

(a) the correlation coefficients calculated using the grade six data were based on a smaller number of pupils, i.e. 429, as compared with the 520 students for whom grade three teacher ratings were available; and

(b) the lapse of three years between the administration of the MAT and the grade three teacher ratings and the administration of the grade six Teacher Rating Questionnaire would likely account for some decrease in these correlation coefficients. Indeed, the decrease is remarkably small given a time interval of three years.

To summarize, there is a relationship between the average MAT scores and the scores for the various subsections of the Teacher Rating Questionnaire. This relationship was strongest for the teacher ratings obtained in grade three (particularly the performance and prediction subsections), the same year that the MAT was administered. Further, the

results indicated that the accuracy of long-range predictions (i.e. grade six) would be slightly less than that of short term predictions. One could hypothesize that one possible reason for the discrepancy between average MAT scores and teacher ratings in both grades three and six is that the class norms on which the teacher ratings are based would usually include a more restricted range of pupil abilities than the standardized norms developed for the MAT and based on a nation-wide sample of pupils.

Prediction of Grade Six Teacher Ratings from Grade Three Teacher Ratings

To what extent do evaluations of pupils made in one grade agree with evaluations made in a later grade? Evaluations made of a pupil's school performance and adjustment tend to follow that pupil throughout his school career in some form of permanent record. When a pupil enters a new grade, the new teacher will likely consult the O.S.R.[5] with the intention of gaining background information and information about previous school success. The ratings are, however, made near the end of the school year: they are made by different teachers and the classmates of the pupils in grade six are unlikely to have been their classmates in grade three. Therefore some disagreement between the ratings of one year and the next could be expected. Although the teacher ratings obtained for the Study of Achievement do not appear on the pupil's permanent school record, they were considered to be similar to the descriptive comments a teacher might include on the O.S.R.

Correlations were obtained between the corresponding subsections of the grade three and the grade six teacher ratings and are presented in Table 3. In addition to the correlation coefficients, the per cent of variability in the data that is explained by each coefficient is shown.

---

5 Ontario School Record

These values (i.e. r^2) represent the accuracy of the prediction that can be made of the grade six teacher ratings from a knowledge of the grade three ratings.

TABLE 3

CORRELATIONS BETWEEN THE CORRESPONDING SUBSECTIONS OF THE GRADE THREE AND GRADE SIX TEACHER RATING QUESTIONNAIRES

Subsections Correlated	r	r^2
Adjustment 3 x Adjustment 6	.54	.29
Performance 3 x Performance 6	.57	.32
Creativity 3 x Creativity 6	.41	.17
Prediction 3 x Prediction 6	.53	.27
Total 3 x Total 6	.60	.36

(N = 429)

It is obvious that the best prediction (36%) could be made when the total scores on the questionnaires were employed. The performance, adjustment and prediction subsections allowed 32, 29 and 27 per cent accuracy respectively, while creativity allowed only 17 per cent. This moderate agreement between scores for grade three and scores for grade six does not suggest either that some teacher ratings were "wrong," or that teacher ratings are therefore poor instruments for pupil evaluation. Rather, they do imply that predictions from one grade to another must be made with some caution. Differences in the two sets of ratings could be due to at least two factors:

(a) changes in the pupils' behaviour during the three years intervening between the two sets of ratings; and/or

(b) the fact that the ratings in grade three and grade six were completed by two different sets of teachers who may have applied different criteria.

It is likely that the adequacy of the frames of reference that teachers have at hand varies with the type of behaviour to be rated. More clearly defined standards are available for judging academic performance than creativity, due to the nature of the school curriculum. This would account for the lowest correlations being obtained for ratings of creativity and highest for performance. The values of the correlations serve as an indication of the extent to which teachers are in agreement as to the criteria or standards to be used in rating pupil behaviours.

Comparison of Teacher Rating Scores and Discrepancy Scores

For each of the TR subsections, mean scores were calculated for each of the five discrepancy score achievement groups. These calculations, conducted separately for boys and girls, are presented in Table 4 and illustrated graphically in Figures 2A and 2B.

An examination of the mean scores for both boys and girls in both grades showed that for all of the TR subsections, the teachers rated over-achievers, i.e. Groups 4 and 5, higher than they rated average-achievers, i.e. Group 3. In addition, both over- and average-achievers were rated higher than under-achievers, i.e. Groups 1 and 2. This relationship was most pronounced for the performance and prediction subsections as is evidenced by the steeper slopes of the lines depicting the relationship between these two subsections and the discrepancy scores, as seen in Figures 2A and 2B for grades three and six respectively. Whether or not a pupil is an over-, average-, or under-achiever is most strongly reflected in teachers' ratings of his academic performance and to a considerably lesser degree in estimates of his general classroom behaviour. It will be remembered that this same pattern of results,

TABLE 4

MEAN TEACHER RATING SUBSECTION SCORES FOR EACH OF THE FIVE DISCREPANCY SCORE ACHIEVEMENT GROUPS, FOR BOYS AND GIRLS, RESPECTIVELY

Teacher Ratings	Achievement Groups				
	1	2	3	4	5
Grade 3					
Adjustment - Male	3.95	4.46	4.65	5.59	5.87
- Female	4.13	4.57	5.55	5.85	6.70
Performance - Male	2.80	3.71	4.44	5.45	5.83
- Female	2.84	3.59	4.88	5.45	5.83
Creativity - Male	3.35	3.54	4.36	5.42	5.30
- Female	3.18	3.59	4.21	4.55	4.64
Prediction - Male	2.55	3.33	4.39	5.42	5.57
- Female	2.18	3.08	4.08	4.75	5.30
Grade 6					
Adjustment - Male	3.40	4.79	4.61	5.00	5.69
- Female	4.03	5.06	5.05	5.51	6.00
Performance - Male	2.91	4.00	4.20	5.03	5.89
- Female	3.36	3.96	4.36	5.18	5.78
Creativity - Male	3.41	4.10	4.17	5.15	5.50
- Female	3.28	3.59	3.88	4.76	5.28
Prediction - Male	2.34	3.75	4.27	5.45	6.11
- Female	2.89	3.64	3.90	5.12	5.78

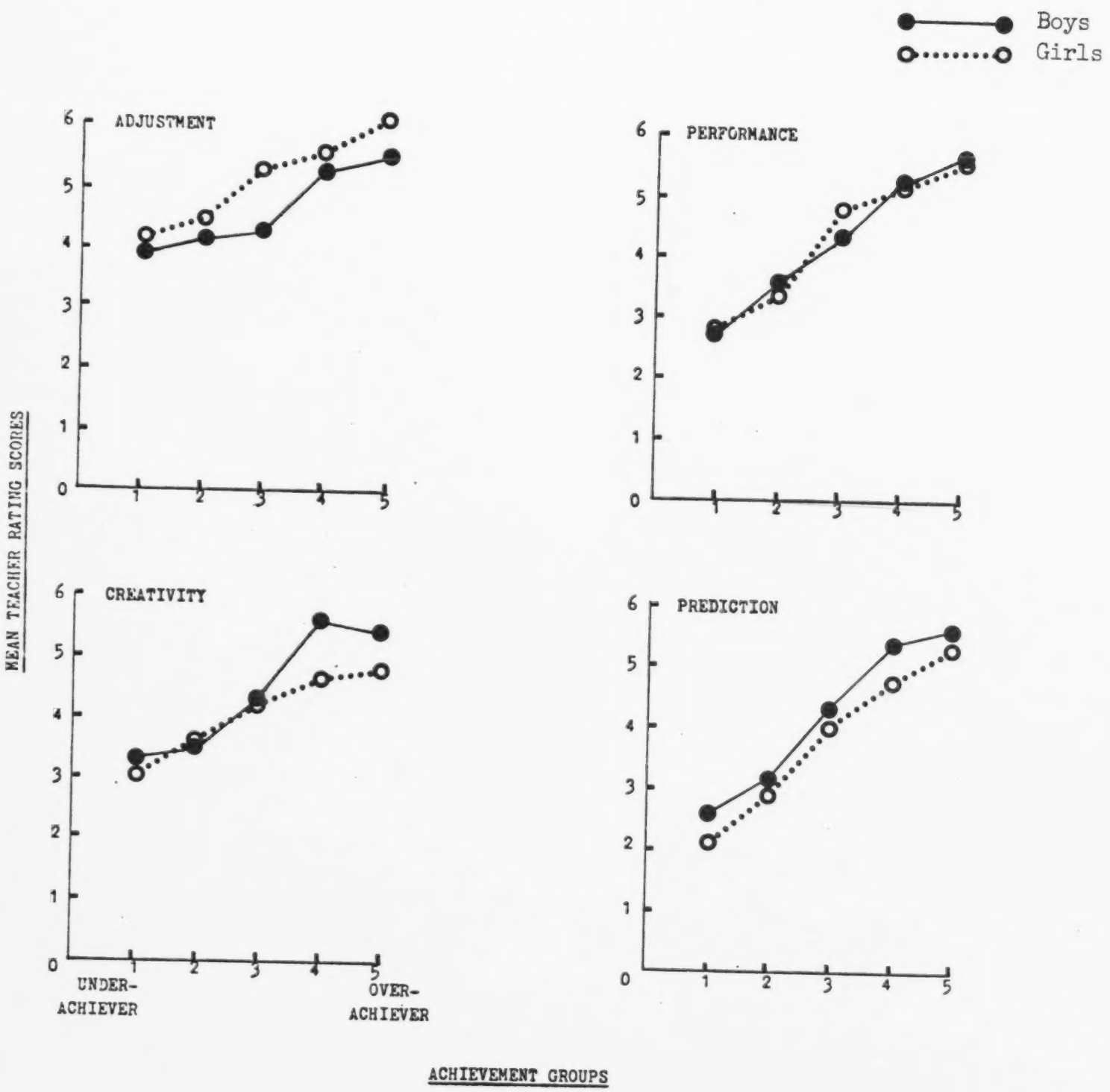

Fig. 2A. Mean scores for each subsection of the grade three Teacher Rating Questionnaire, calculated separately for boys and girls in each of the five achievement groups.

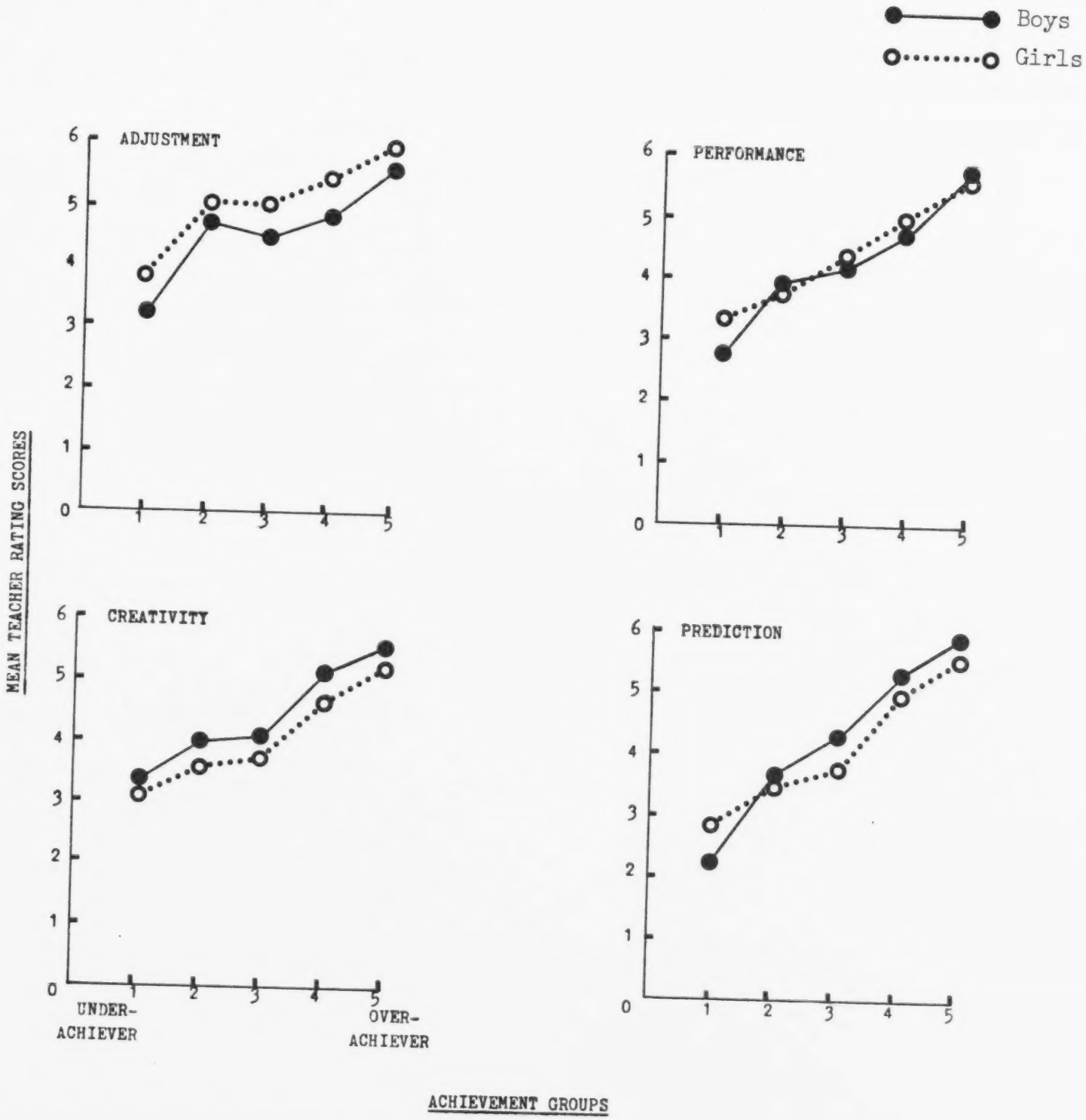

Fig. 2B. Mean scores for each subsection of the grade six Teacher Rating Questionnaire, calculated separately for boys and girls in each of the five achievement groups.

as discussed previously, characterized the relationship between the TR subsection scores and the average MAT scores.

Although there was little change in the mean ratings between grades, there were differences within each discrepant achievement group between mean scores obtained for girls and those obtained for boys. The mean ratings were higher for girls than for boys on the adjustment subsection, and higher for boys than for girls on the creativity and prediction subsections. Boys and girls were <u>not</u> rated differently on the performance subsection. Since boys were rated as likely to go further in school, although their academic performance was rated as being similar to that of the girls, it seems evident that factors other than academic performance alone were influencing the teachers' predictions of how far their pupils would go in school. The most probable factor is a still existing expectation in our society that boys, as a group, will obtain more education than girls. Given equal intellectual abilities, boys will more likely be encouraged to continue their schooling, and in turn, it is likely that teachers, on the average, responded accordingly in their ratings.

The Relationship Between I.Q. Scores and Average MAT Scores and Teacher Rating Scores

To provide a measure of the degree of relationship between I.Q. scores and each of the achievement measures, i.e. MAT scores and teacher ratings, correlation coefficients were calculated (see Table 5).

TABLE 5

CORRELATIONS OF I.Q. WITH MEAN MAT SCORES AND TEACHER RATING SUBTEST SCORES

	MAT	Adjustment	Performance	Creativity	Prediction	Total
Grade 3						
r	.54	.25	.40	.37	.48	.41
r^2	.29	.06	.16	.14	.23	.17
Grade 6						
r	--	.25	.39	.32	.45	.39
r^2	--	.06	.15	.10	.20	.15

The correlation obtained between the I.Q. scores and the mean MAT scores was .54. The correlations between I.Q. and each of the four TR subsections and the total score in both grades three and six ranged from .25 to .48; therefore, the performance subsection of the Teacher Rating Questionnaires was a better predictor of the average MAT scores than the I.Q. scores (see Table 2). Although it is evident that a knowledge of a pupil's I.Q. score does not enable one to make a very accurate prediction of either his average MAT score or his TR subsection scores, as noted in a previous paper[6], the pattern of these results is encouraging as

> "...they seem to indicate that the teacher uses the I.Q. information in those instances where it likely has most relevance, i.e. predicting how far the child will go in school and is least influenced by it in her rating of whether the child presents a discipline problem

...continued

6 For a more detailed discussion of the importance of I.Q. as a predictor of a child's school performance, the reader is referred to School Achievement: A Preliminary Look at the Effects of the Home, Research Department, 1970.

> in the classroom. Children with lower I.Q. scores generally were no more likely to be rated as being discipline problems than children with higher I.Q. scores."
>
> (Crawford & Eason, 1970, pp. 18-19)

As was pointed out in School Achievement: A Preliminary Look at the Effects of the Home, although I.Q. was a better predictor of both the MAT scores and the TR scores than either socio-economic status or a sample of home environment factors, it could explain a maximum of only 30% of the variability in these achievement measures. There are undoubtedly factors other than the pupil's measured intelligence that strongly influence his school performance, whether that performance is measured by a standardized test or by teachers' ratings.

MAT Scores, I.Q. Scores and Discrepancy Scores as Combined Predictors of Teacher Ratings

The correlations obtained between the teacher ratings and each of the standardized measures included in this study, i.e. I.Q. scores, MAT scores and discrepancy scores, indicated that some relationship exists between these measures. The question arose, was there a way of comparing all three standardized measures simultaneously to determine if there was a combination of the measures which would increase the accuracy of predicting the teacher ratings as compared with that which could be obtained using any one measure alone? One way of accomplishing this goal was to utilize the statistical technique of multiple regression analysis in which the teacher ratings were designated as the criterion variables and I.Q., MAT scores and discrepancy scores were the predictor variables.

Multiple regression analysis is a complex form of correlation analysis that proportions variables on the basis of their correlation with the criterion variable in such a way as to provide the best possible,

i.e. maximum, prediction of the criterion variable. An optimal proportioning of the predictor variables is accomplished by assigning weights to all the variables so that the most efficient predictor receives the largest weight. Multiple regression analysis selects from among the predictor variables only those variables that add to the accuracy of the prediction. Eventually, a point will be reached where the addition of further variables does not increase the accuracy of the prediction. At this point, a numerical value is obtained indicating how much of the variability in the data has been accounted for by the selected predictor variables. Thus multiple regression analysis selects the minimum number of variables that give a maximum amount of prediction. When all the predictor variables originally entered into the regression are good predictors, the minimum number of variables selected as contributing to the prediction of the criterion variables will be the same as the total number of predictor variables originally entered.

Interpretation of a multiple regression analysis is similar to the interpretation of a simple correlation. The value of R obtained by multiple regression analysis is simply the correlation between the criterion variable and the sum of the weighted predictor variables. R^2 is an indication of the amount of variation in the data that is accounted for by the selected set of predictor variables.

Using the scores for each subsection and the total score of the teacher ratings as the criterion variables, and the average MAT scores, I.Q. scores and discrepancy scores as the predictor variables, two sets of multiple regression analyses were conducted, one for each of grade three and grade six. The results of these analyses are summarized in Table 6.

TABLE 6

VALUES OBTAINED FOR EXPRESSING THE DIFFERENT MULTIPLE REGRESSION EQUATIONS FOR EACH OF THE TEACHER RATING SUBSECTIONS IN BOTH GRADE THREE AND GRADE SIX WHEN ACHIEVEMENT GROUPS, AVERAGE MAT SCORES AND I.Q. WERE THE PREDICTOR VARIABLES

Criterion	R	R^2	Predictor Variables*			Regression Constant
			Achievement Group	I.Q.	MAT	
Grade 3						
Adjustment	.4976	.2476	.0371	--	.4673	2.5608
Performance	.6899	.4759	.1061	.1086	.5447	-11.2330
Creativity	.5146	.2648	--	.1347	.4473	-4.6489
Prediction	.6842	.4682	.0614	.2071	.5083	-4.5362
TOTAL	.6658	.4433	.0483	.0964	.5727	-17.6299
Grade 6						
Adjustment	.4264	.1818	.1439	.1353	.2370	2.8092
Performance	.6233	.3885	--	.1279	.5217	-10.1022
Creativity	.4680	.2190	.0525	.1584	.3322	-3.1146
Prediction	.6620	.4383	--	.1735	.5395	-4.8525
TOTAL	.6103	.3724	.0829	.1605	.4540	-15.3324

* The values listed for each of the three predictor variables are the beta weights obtained in the regression analysis.

The accuracy of prediction obtained for each of the teacher rating subsections for grades three and six respectively was: adjustment, 25% and 18%; performance, 48% and 39%; creativity, 26% and 22%; prediction of school success, 47% and 44%; total score, 44% and 37%.

A comparison of these values of R^2 obtained from the multiple regression analyses, with the corresponding values of r^2 for the correlations between each of the grade three and grade six teacher ratings and the average MAT scores (see Table 2), indicated that very little was added to the predictability of teacher ratings when all three standardized measures were combined.

Further confirmation of the importance of the role of the pupils' work performance to the teachers' ratings comes from an examination of the beta weights[7] of the multiple regression analyses. In each of the ten analyses, the highest weight was assigned to the average MAT scores, meaning that these scores contributed the most to the prediction of each of the TR subsection scores. The contribution of the I.Q. and discrepancy scores was approximately one-half or less of the contribution of the MAT scores.

---

7 These are the weightings which, as described on page 29, permit maximum predictability through the optimal proportioning of the predictor variables.

SUMMARY AND IMPLICATIONS

In this report, teacher ratings were compared with standardized measures of achievement to determine the extent of their relationship. The question of interest was, does a teacher's rating of a pupil in her class provide the same information about his level of achievement as scores he obtains on a standardized test, the Metropolitan Achievement Test?

Quite bluntly, the answer to this question is no! Although there is some relationship between these two measures, the degree of relationship depends on those areas in which we ask the teacher to make ratings! An examination of the data obtained in both grade three and grade six revealed a consistent pattern in the relationship between various subsections of the Teacher Rating Questionnaire and the MAT, the discrepancy scores and I.Q. scores. Teachers' ratings of pupils' performance was more closely related to these measures than were their ratings of pupils' adjustment and creativity.

Similarly, teachers in grade three and grade six were in greater agreement about their ratings of performance and prediction than their ratings of classroom behaviour and creativity. Furthermore, the teachers' ratings of performance provided a better prediction of the achievement test scores than did the I.Q. test scores.

It appears then that teachers' ratings are good predictors in certain areas. In spite of differences among teachers and among classes, as well as the development in the pupils' characteristics from grade three to grade six, there was a reasonable level of agreement among the teachers in their ratings of students performance. Teacher differences in

definitions of what constituted a particular type of behaviour were more apparent in the ratings of classroom behaviour, and, especially, of creativity. A lack of some clear, generally available, criteria applicable to these areas is evident. Teachers' judgements of "high" and "low" achievement were consistent with the achievement test results. This suggests that the teachers do indeed use classroom performance as a criterion in making their ratings much more than any knowledge of a recorded I.Q. score. This is supported by the regression analyses.

In conclusion, the data suggest that teachers' ratings of academic performance are an important and dependable adjunct to the use of standardized achievement tests.

REFERENCES

Alexander, A. M. Teacher judgement of pupil intelligence and achievement is not enough. Elementary School Journal, 1953, 53, pp. 396-401.

Crawford, Patricia, & Eason, G. School achievement: a preliminary look at the effects of the home. Toronto: The Board of Education for the City of Toronto, Research Department, 1970 (#83).

Ebbesen, J. Kindergarten teacher ratings as predictors of academic achievement in the primary grades. Journal of Educational Measurement, 1968, 5, pp. 259-262.

Finley, Carmen J. How well can teachers judge achievement? The case of the illusive criterion. California Journal of Educational Research, 1966, 17, pp. 126-132.

Haberman, M., & Raths, J. D. High, average, low -- and what makes teachers think so. Elementary School Journal, 1968, 68, pp. 241-245.

Hubbard, R. E., & Flesher, W. R. Intelligent teachers and intelligence tests -- do they agree? Educational Research Bulletin, 1953, 32, pp. 113-122 and 139-140.

Meyers, C. E., Attwell, A. A., & Orpet, R. E. Prediction of fifth grade achievement from kindergarten test and rating data. Educational and Psychological Measurement, 1968, 28, pp. 457-463.

Thorndike, R. L. The concepts of over- and underachievement. New York: Bureau of Publications, Teacher's College, Columbia University, 1963.

APPENDIX

STUDY OF ACHIEVEMENT

STAGE V

ADMINISTRATION BOOKLET

TEACHERS' RATING QUESTIONNAIRE

(GRADE 3)

INSTRUCTIONS

1. Please read each question carefully.
2. Decide from your own knowledge the ratings for each child.

ADJUSTMENT TO SCHOOL PROGRAMME

1. Discipline

Displays behaviour that you, the teacher, consider appropriate for your classroom.

Rate 0: Constant discipline problem; behaviour always inappropriate.

2: Frequent discipline problem; behaviour often inappropriate.

4: Occasional discipline problem; exercises some self control.

6: Very seldom causes discipline problems, exercises self control most of the time.

8: Never causes discipline problems, behaviour always appropriate.

2. Acceptance of Routines

Accepts responsibility in connection with classroom work, seatwork, routine rules and directions.

Rate 0: Never accepts responsibility; needs constant help and attention from teacher.

2: Seldom accepts responsibility; has to be coaxed, inconsistent in his response to routines.

4: Frequently accepts responsibility; tries to look after his tasks.

6: Regularly accepts responsibility; looks after his tasks almost always.

8: Consistently accepts responsibility; looks after his tasks successfully all the time.

3. Acceptance of Goals

Shows desire to make positive contributions towards classroom activities, i.e., answers questions readily, moves in gym willingly, sings in music periods, talks during discussions.

Rate 0: Shows no interest in the activities, makes no contribution.

2: Shows limited interest in a few activities.

4: Responsive towards numerous activities, able to contribute sometimes.

6: Shows interest in a great number of activities, contributes often.

8: Is interested in all activities and contributes whenever possible.

4. Ability to Get Along

Interacts with most of his classmates in a satisfactory manner.

Rate 0: Unable to get along in classroom, (or in schoolyard), always quarrelsome in social contacts.

2: Frequently quarrelsome, or limits social contacts to one or two chosen friends.

4: Gets along with most pupils, and regularly participates in group activities.

6: Often shows leadership ability in group activities, and is popular with most classmates.

8: Consistently shows leadership ability in social contacts, and is trusted by other children.

WORK PERFORMANCE

5. Attention and Work Completion

Has good attention span, is able to attend to teacher and assignments without constant encouragement.

Rate 0: Extremely short attention span, easily distracted, seldom if ever, finished assignments.

2: Short attention span, easily distracted, gets work done occasionally.

4: Able to listen for the duration of the lesson, usually gets his work done.

6: Above average attention span, gets his work done regularly.

8: Superior attention span, will work at any task as long as necessary, till it is completed.

6. Reading

Reads with comprehension and fluency, conveys meaning to listeners.

Rate 0: Reads with little or no comprehension, mostly word by word, without much meaning.

2: Reads with word recognition and comprehension at bottom level of class.

4: Reads with comprehension and fluency, conveys meaning at middle level of class.

6: Reads with word recognition and comprehension at top level of class.

8: Superior reader, able to comprehend most material encountered, e.g., magazines and books at higher grade levels.

7. Language, Self Expression

Can tell or write "stories"; expresses self clearly.

Rate 0: -- Occasionally attempts to tell or write a "story";
-- "Stories" consist of one or two sentences;
-- Sentences may be completely unrelated.

2: -- Frequently attempts to tell, or write, a "story";
-- "Stories" have many irrelevant ideas.

4: -- Regularly attempts to tell or write a "story";
-- Few, if any, irrelevant ideas.

6: -- Consistently attempts to tell or write "stories";
-- Few, if any, irrelevant ideas;
-- Occasional use of complex sentences.

8: -- Tells or writes coherent "stories";
-- No irrelevant ideas, use of complex and compound sentences;
-- Unusually good command of language.

8. Accuracy and Quality of Work

Can usually do work correctly.

Rate 0: Consistently makes errors in copying and seldom, if ever, does assignments the right way.

2: Inconsistent both in accuracy of copying, and in doing assignments.

4: Does work the right way, but needs supervision.

6: Does work the right way and seldom makes errors.

8: Work always accurate with quality beyond requirements.

9. Mathematical Ability

Shows understanding of mathematical concepts and operations, can solve problems.

Rate 0: Very limited ability to understand mathematical concepts and operations, cannot solve problems.

2: Mathematical understanding and problem solving ability is at lower level of class.

4: Usually able to understand mathematical concepts and operations when presented by teacher.

6: Mathematical understanding and problem solving ability is at upper level of class.

8: Superior mathematical ability, - immediately understands mathematical ideas presented by teacher.

CREATIVE THINKING

FOR YOUR GUIDANCE - the following meanings are intended when these words are used.

Intuition: -- Immediate insight;
-- Immediate apprehension by the mind without reasoning.

Divergent: -- Capable of going in different directions;
-- Differ from the usual.

Inventiveness: -- Ability to devise, to originate.

Imagination: -- Mental faculty of forming images of external objects not present to the senses.

10. Imagination and Inventiveness

Regardless of academic achievement, he may be considered imaginative and inventive.

Rate 0: Never shows imagination or inventiveness.

2: Rarely shows imagination or inventiveness.

4: Occasionally shows imagination or inventiveness.

6: Frequently shows imagination or inventiveness.

8: Regularly shows imagination or inventiveness.

11. Use of out-of-school experiences in class

Draws on background experiences.

Rate 0: Shows no background experiences, reports no information pertaining to the world about him.

2: Shows a few background experiences, reports some information pertaining to the world about him.

4: Reasonably well informed.

6: As a result of his background experiences, he is often able to contribute new information.

8: As a result of his background experiences, regularly displays a wealth of knowledge. High degree of sensitivity to the world around him.

12. Creativity

Shows an urge to explore and create; is intuitive.

Rate 0: Always placid, never shows signs of curiosity, no capacity to be "disturbed."

2: Rarely shows curiosity or the desire to explore.

4: Occasionally displays signs of divergent thinking.

6: Frequently displays signs of divergent thinking, has a great urge to explore.

8: Regularly displays signs of divergent thinking, possesses the rare gift of immediate insight.

SCHOOL ABILITY

13. To provide your estimate of this child's ability, try to predict how far you think he will go (ignore financial ability of parents).

Rate 0: Will have difficulty completing grade eight.

2: Will not complete high school.

4: Will complete high school.

6: Will go to university.

8: Will go beyond a B.A.

STUDY OF ACHIEVEMENT

STAGE VIII

ADMINISTRATION BOOKLET

TEACHERS' RATING QUESTIONNAIRE

(Grade 6)

May, 1968

Instructions

1. Please read each question carefully.
2. Decide from your own knowledge the ratings for each pupil.
3. Use the special pencil to mark the rating on the pupil's I.B.M. card. (Please check that the column used matches the item being marked.) One mark and one mark only per column. As there are only twelve questions, you only have to mark the first four columns.

ADJUSTMENT

1. Discipline

Displays behaviour that you, the teacher, consider appropriate, for your classroom.

Rate 0 - Constant discipline problem; behaviour always inappropriate.

Rate 2 - Frequent discipline problem; behaviour often inappropriate.

Rate 4 - Occasional discipline problem; exercises some self control.

Rate 6 - Very seldom causes discipline problems; exercises self control most of the time.

Rate 8 - Never causes discipline problems; behaviour always appropriate.

2. Ability to Get Along

Interacts with most of his classmates in a satisfactory manner.

Rate 0 - Never able to get along in classroom (or in schoolyard).

Rate 2 - Frequently quarrelsome, or limits social contacts to one or two chosen friends.

Rate 4 - Gets along with most pupils, and regularly participates in group activities.

Rate 6 - Often shows leadership ability in group activities, and is popular with most classmates.

Rate 8 - Consistently shows leadership ability in social contacts, and is trusted by other children.

3. Acceptance of Goals

Contributes to classroom activities, i.e., answers questions readily, talks during discussion, makes active contribution to class projects.

Rate 0 - Shows no interest; makes no contribution.

Rate 2 - Shows limited interest in a few activities.

Rate 4 - General interest in classroom activities; contributes occasionally.

Rate 6 - Wide variety of interests; contributes regularly.

Rate 8 - Participates actively and enthusiastically in all activities; contributes more than do most other pupils.

4. General Adjustment Evaluation

Considering all aspects of the child's adjustment to the classroom environment, evaluate his position.

Rate 0 - Quite out-of-place.

Rate 2 - Seems uncomfortable.

Rate 4 - Adequately adjusted.

Rate 6 - Makes consistent, conscientious efforts to improve himself in relation to his school world.

Rate 8 - Makes an effort to positively influence his classroom world.

PERFORMANCE

5. Reading

Reads with comprehension and fluency; conveys meaning to listeners.

Rate 0 - Reads with little or no comprehension, mostly word by word, without much meaning.

Rate 2 - Reads with word recognition and comprehension at bottom level of class.

Rate 4 - Reads with comprehension and fluency; conveys meaning at middle level of class.

Rate 6 - Reads with word recognition and comprehension at top level of class.

Rate 8 - Superior reader, able to comprehend most material encountered, e.g., magazines and books at higher grade levels.

6. Mathematical Ability

Shows understanding of mathematical concepts and operations; can solve problems.

Rate 0 - Very limited ability to understand mathematical concepts and operations; cannot solve problems.

Rate 2 - Mathematical understanding and problem solving ability is at lower level of class.

Rate 4 - Usually able to understand mathematical concepts and operations when presented by teacher.

Rate 6 - Mathematical understanding and problem solving ability is at upper level of class.

Rate 8 - Superior mathematical ability; immediately understands mathematical ideas presented by teacher.

7. Language

Extent of vocabulary; correct grammatical usage of English; ability to express self clearly. (Both oral and written.)

Rate 0 - Generally very poor command of the language, frequently misunderstood, or cannot express self due to inadequate language.

Rate 2 - Language poor; on occasion is misunderstood.

Rate 4 - Command of language adequate; child still makes some grammatical errors.

Rate 6 - Good command of language; no grammatical mistakes.

Rate 8 - Extremely articulate for child this age; superior vocabulary, clear expression of ideas.

8. Use of Out-of-School Experiences in Class

Draws on background experiences, reading.

Rate 0 - Shows no background experiences; reports no information pertaining to the world about him.

Rate 2 - Shows a few background experiences; reports some information.

Rate 4 - Reasonably well informed.

Rate 6 - As a result of his background experiences, he is often able to contribute new information.

Rate 8 - As a result of his background experiences, regularly displays a wealth of knowledge. High degree of sensitivity to the world around him.

9. General Performance Level

The quality of work; diligence in performing it.

Rate 0 - Makes many errors; doesn't concentrate; seems uninterested in improvement.

Rate 2 - Makes an effort to concentrate, still has difficulty with work; quite a few errors.

Rate 4 - Listens; performs as required; relatively neat.

Rate 6 - Somewhat above average; diligent; few errors.

Rate 8 - Far above average; diligent; produces extremely accurate work.

CREATIVE THINKING

For Your Guidance -- the following meanings are intended when these words are used:

Intuition - immediate insight;
- immediate apprehension by the mind without reasoning.

Divergent - capable of going in different directions;
- differ from the usual.

Inventiveness - ability to devise, or originate.

Imagination - mental faculty of forming images of external objects not present to the senses.

10. Imagination and Inventiveness

Regardless of academic achievement, he may be considered imaginative and inventive.

Rate 0 - Never shows imagination or inventiveness.

Rate 2 - Rarely shows imagination or inventiveness.

Rate 4 - Occasionally shows imagination or inventiveness.

Rate 6 - Frequently shows imagination or inventiveness.

Rate 8 - Regularly shows imagination or inventiveness.

11. Creativity

Shows an urge to explore and create; is intuitive.

Rate 0 - Always placid, never shows signs of curiosity; no capacity to be "disturbed."

Rate 2 - Rarely shows curiosity or the desire to explore.

Rate 4 - Occasionally displays signs of divergent thinking.

Rate 6 - Frequently displays signs of divergent thinking; has a great urge to explore.

Rate 8 - Regularly displays signs of divergent thinking; possesses the rare gift of immediate insight.

ABILITY

12. School Ability

To provide your estimate of this child's ability, try to predict how far you think he will go (ignore financial ability of parents).

Rate 0 - Will have difficulty completing Grade Eight.

Rate 2 - Will not complete high school.

Rate 4 - Will complete high school.

Rate 6 - Will go to university.

Rate 8 - Will go beyond a B.A.

A Brief Chronology of the Development of the Teacher Rating Questionnaires

When the Study of Achievement was initiated, it was believed that the teacher's assessment of her pupils was one important indication of achievement that must be included. In consultation with the Kindergarten Department, five areas of importance in terms of the kindergarten programme were identified: language, mental, social, emotional and physical development. Mrs. Helen Gaston,[8] who was then a kindergarten teacher, made a major contribution in the development of the first form of the rating scale (Stage II). She designated 40 behaviourally defined situations for which the teacher was to provide a rating of each pupil. Pilot work showed that the teachers preferred a five point rating scale as follows: 0, 2, 4, 6, 8. On the basis of her experience, Mrs. Gaston developed descriptive statements that defined (a) typical or average behaviours, to be rated as 4, (b) extreme behaviours to be rated as 0 or 8 depending upon the direction of the behaviour displayed, and (c) two intermediate categories to be rated as 2 or 6.

The next year, it was apparent that the rating scale for grade one (Stage III) required modifications because some behaviours considered to be typical in grade one differed from those considered to be typical in kindergarten. The number of items included in this scale was slightly reduced (i.e. 33 items as compared to 40 items) and a more elaborate behavioural description provided for each item. Miss Elinor Gullette, a consultant, did the work on Stage III.

In this same year, the Kindergarten Questionnaire was revised by Mrs. Gaston so that the instructions were clearer. It was made available to kindergarten teachers as a device to be used at the end of the year to assess

---

8 Mrs. Gaston is currently Kindergarten Supervisor for the Board of Education for the Borough of Etobicoke.

the achievement of their students. In this revised form it was called the "June Questionnaire" and contained 39 items. An additional modification was made by Mrs. Gaston so that a similar form dealing with the same topics could be used by teachers early in the school year to consider school progress; this is referred to as the "Fall Questionnaire," (24 items). Both questionnaires are available to kindergarten teachers from the Kindergarten Department.

In grade two (Stage IV) a further modification, carried out by a consultant, Miss Florence Roliff, reduced the rating scale to 27 items. Although there was some variation in content, a serious attempt was made to define similar behavioural situations for all three grades. Each year, comments by teachers provided assistance in making modifications and in directing the Department to attempt a shorter, more streamlined questionnaire.

For grade three (Stage V) Mrs. Szabo, a vice-principal at that time, now a principal, assisted in the questionnaire modifications. As a result of previous experience, discussions, and primarily her suggestions, modifications were undertaken which resulted in the abbreviated form analyzed in this report. This questionnaire now had 13 items which provided ratings in the areas of Adjustment to School Programme, Work Performance, Creativity and Prediction of School Success. No further major revisions were made to this questionnaire although some of the items were modified and one item from the section on creativity was dropped in grade four (Stage VI). In grade six (Stage VIII), the questionnaire was identical to the one used in grade four. In this form, the questionnaire seemed so successful that it was used as one of the instruments in the Study of New Canadians. For this purpose it was considered necessary to drop the two items on creativity as well as the item on mathematics since in some

instances the ratings were being completed by teachers who did not have the opportunity to assess the behaviours mentioned in these three items.

The most noteworthy difference between the form used from grade three onwards, and the form used from kindergarten to grade two is that in the rating scales for the early grades the teachers were asked to rate each child in specific situations, e.g., can follow some directions in games, shows evidence of eye-hand coordination when building with blocks, has ability to control temper. In each of these situations, specific behavioural examples were provided for each rating level. In the later grades, the situations rated were much more general in nature, e.g., can accept responsibility in the class, gets along with most of his classmates, can read fluently etc. The typical behaviour <u>for the grade</u> was used as the referent for each of these situations.

The Teacher Rating Questionnaire is now available to be used at any grade level from kindergarten to high school; similar means and standard deviations have been found for these questionnaires when used at different grade levels.